



Random Forests on Hierarchical Multi-Scale Supervoxels for Liver Tumor Segmentation in Dynamic Contrast-Enhanced CT Scans.

Pierre-Henri Conze, Vincent Noblet, François Rousseau, Fabrice Heitz,
Riccardo Memeo, Patrick Pessaux

► To cite this version:

Pierre-Henri Conze, Vincent Noblet, François Rousseau, Fabrice Heitz, Riccardo Memeo, et al.. Random Forests on Hierarchical Multi-Scale Supervoxels for Liver Tumor Segmentation in Dynamic Contrast-Enhanced CT Scans.. International Symposium on Biomedical Imaging, Apr 2016, Prague, Czech Republic. hal-01284631

HAL Id: hal-01284631

<https://hal.science/hal-01284631>

Submitted on 7 Mar 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RANDOM FORESTS ON HIERARCHICAL MULTI-SCALE SUPERVOXELS FOR LIVER TUMOR SEGMENTATION IN DYNAMIC CONTRAST-ENHANCED CT SCANS

P.-H. Conze^{}, V. Noblet^{*}, F. Rousseau[◇], F. Heitz^{*}, R. Memeo[△] and P. Pessaux[△]*

^{*} ICube, Université de Strasbourg, CNRS, FMTS, Strasbourg, France

[◇] Institut Mines-Télécom, Télécom Bretagne, INSERM, LATIM, Brest, France

[△] Institut Hospitalo-Universitaire de Strasbourg, France

ABSTRACT

This paper addresses multi-label tissue classification in the context of liver tumor segmentation for patients with hepatocellular carcinoma (HCC). Covering such issue in an interactive perspective through supervoxel-based random forest (RF) requires an adaptive data sampling scheme to deal with multiple spatial extents and appearance heterogeneity. We propose a simple and efficient strategy combining standard RF and hierarchical multi-scale tree resulting from recursive 3D SLIC supervoxel decomposition. By concatenating features across the hierarchical multi-scale tree to describe leaf supervoxels, we enable RF to automatically infer the most informative scales discriminating tissues based on their intrinsic properties. Our method does not require any explicit rules on how to combine the different scales. Quantitative assessment on expert ground truth annotations demonstrates improved results compared to standard single-scale strategies for HCC tumor segmentation in dynamic contrast-enhanced CT scans.

Index Terms— liver tumor segmentation, random forest, supervoxels, hierarchical multi-scale tree, spatial adaptivity

1. INTRODUCTION

We tackle the challenging problem of multi-label tissue classification towards accurate tumor segmentation. Classifying different tissue types remains an open issue due to wide appearance heterogeneity, severe class overlap in feature space, ambiguous boundaries, high diversity in shape, location and size as well as large data volumes. Tissue-specific segmentation covered in an interactive perspective through voxel-wise supervised ensemble learning has grown in popularity, especially using random forests (RF) [1]. One key aspect of recent works is how to capture long-range spatial context using dedicated context-rich features [2], entangled decision forests [3] or conditional random fields regularization [4]. Alternatively, [5, 6] extend the use of RF from voxels to supervoxels. While ensuring weak interaction efforts, it improves sample

representativeness, allows to regularize features over consistent areas and reduces computational complexity.

However, these single-scale strategies are limited in their ability to deal with spatial adaptivity. Spatial extent of two different structures may highly differ and intrinsic tissue properties may emerge only at some specific scales. Multi-scale image and feature representations are thus often better suited. In [7], the most informative scale to compute features is inferred during training. [8] relies on a hierarchical multi-scale supervoxel representation and learns the optimal image sampling in an extended RF framework. Nevertheless, it requires to define explicit rules to go through one scale to another.

In this work, we introduce a different, simple and efficient strategy combining standard RF and hierarchical multi-scale tree resulting from recursive supervoxel decomposition. Our motivation is to describe each leaf supervoxel as a sequence of supervoxels belonging to its ascendant hierarchy. By concatenating features across the hierarchical multi-scale tree, we enable RF to find itself the best data sampling without explicitly choosing how to combine the different scales.

This study is motivated but not limited to the clinical management of hepato-cellular carcinoma (HCC), the most common type of liver cancer, which requires accurate HCC tumor segmentation for computed-aided diagnosis [9]. It translates in classifying liver into parenchyma, tumoral active and necrotic areas (Fig.1b) using dynamic contrast-enhanced (DCE) CT scans. This work extends [6] from single to multi-scale supervoxel-based RF classification and reaches a further step towards more accurate HCC tumor segmentation.

2. METHOD

2.1. Multi-phase input data

Tumor segmentation benefits in our context from DCE-CT scans since practitioners focus on the association of both arterial hypervascularity and venous washout for HCC diagnosis [9]. In practice, a bolus of contrast agent is injected to patients and CT scans are acquired before injection (WI) and after at arterial (AR), early venous (EV) and late venous (LV) phases. Each examination consists in a set of images

This work received the financial support from Fondation Arc, www.fondation-arc.org.

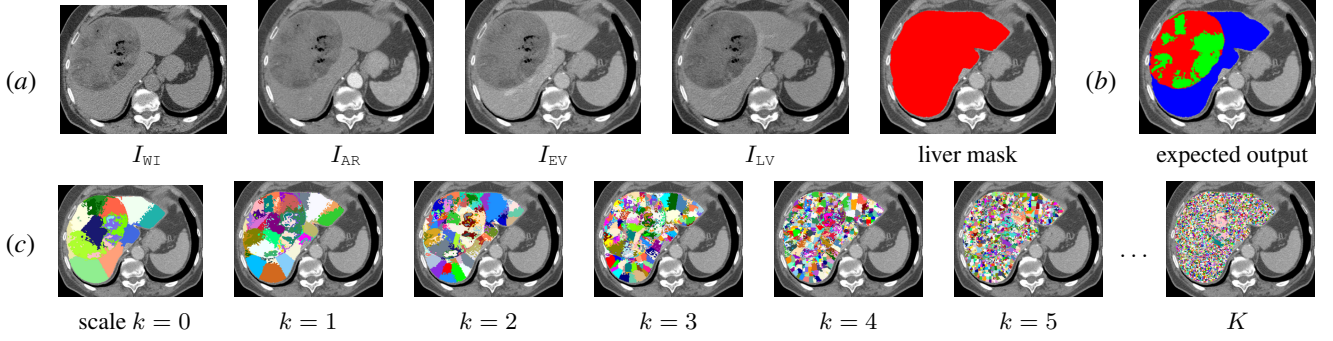


Fig. 1. Registered multi-phase input data obtained from DCE-CT scans with associated liver mask (a). Liver segmentation into parenchyma (blue), tumoral active (red) and necrotic (green) areas (b) performed using hierarchical multi-scale supervoxels (c).

$S_I = \{I_{WI}, I_{AR}, I_{EV}, I_{LV}\}$ registered with respect to I_{EV} since I_{EV} exhibits greater inter-class contrasts than other acquisitions (Fig.1a). In S_I , each image $I_p : \Omega \subset \mathbb{Z}^3 \rightarrow \mathbb{Z}$ associates a greyscale value $I_p(\mathbf{x})$ to each voxel \mathbf{v} located at $\mathbf{x} \in \Omega$ with $p = \{WI, AR, EV, LV\}$. A liver segmentation mask (Fig.1a) is provided at EV phase and defines the liver area Ω_l .

2.2. Hierarchical multi-scale supervoxel representation

Contrary to standard supervoxel decomposition defined on one single scale [5, 6], we rely on a hierarchical multi-scale supervoxel representation similarly to [8]. Such data representation requires a decomposition of the liver area Ω_l into a set of $K + 1$ partitions $\mathcal{P}^k = \{r_i^k\}$ defined at scale $k \in \{0, \dots, K\}$ (Fig.1c) where 0 and K denote respectively the coarsest and finest scales. Each partition $\{\mathcal{P}^k\}$ is a collection of compact 3D supervoxels $\{r_i^k\}$ built at scale k such that $r_i^k \cap r_j^k = \emptyset$ and $\bigcup_i r_i^k = \Omega_l$. Supervoxels are generated using a 3D extension of the simple linear iterative clustering (SLIC) superpixel algorithm [10] which aggregates neighboring voxels based on both spatial and intensity proximity.

The resulting sequence of partitions $\{\mathcal{P}^k\}$ is encoded in the layers of a multi-resolution tree $\mathcal{M} = \{\mathcal{M}^k\}$ where layer \mathcal{M}^k maps each supervoxel $r_i^k \in \mathcal{P}^k$ to a set of child supervoxels $\{r_j^{k+1}\} \subset \mathcal{P}^{k+1}$ (Fig.2) such that $r_i^k = \bigcup_j r_j^{k+1}$.

This representation is obtained through a recursive supervoxel generation process. Instead of merging supervoxels in a fine-to-coarse strategy [8], we first generate an initial coarse partition \mathcal{P}^0 by dividing Ω_l into a small set of large supervoxels through SLIC. Each of these supervoxels r_i^0 of \mathcal{P}^0 is then split into K_s supervoxels $\{r_j^1\} \subset \mathcal{P}^1$ using SLIC applied on the area formed by voxels of r_i^0 . Once built, each r_j^1 is split into K_s children and this procedure is repeated iteratively in a coarse-to-fine way down to the K^{th} layer of \mathcal{M} .

2.3. Supervoxel-based random forest classification

In single-scale supervoxel-based classification [5, 6], the supervoxel resolution has to be carefully selected. A compromise must be reached between large supervoxels which may

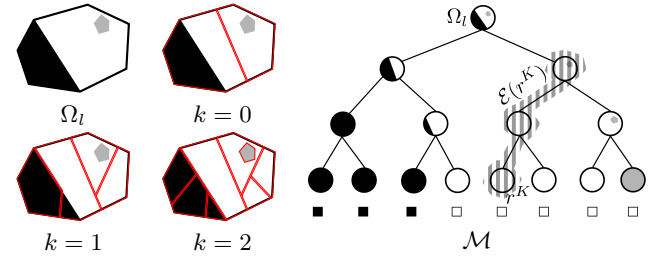


Fig. 2. Illustration of the hierarchical multi-scale supervoxel representation on a binary classification example with $K = 2$ and $K_s = 2$. Expected labels are displayed into squared boxes.

overlap different tissue types and small ones which do not take full advantage of spatial context. Manually setting the scale is a cumbersome task which requires exhaustive trials to get the optimal resolution. Even the best compromise may not provide satisfactory results since spatial adaptivity is required to infer discriminative scale-dependent tissue properties.

In the proposed method, we let a classifier find itself the best data sampling using the hierarchical multi-scale tree \mathcal{M} resulting from recursive supervoxel decomposition (Sec. 2.2). Once \mathcal{M} is built, we start by assigning visual features $\theta(r^k) : \mathcal{P}^k \rightarrow \mathbb{R}$ to all 3D supervoxels r^k in each partition \mathcal{P}^k with $k \in \{0, \dots, K\}$. Then, we associate to each supervoxel r^K at finest scale K all the supervoxels of decreasing scale belonging to its ascendant hierarchy including itself: $\mathcal{E}(r^K) = \{r^k\}_{k \in [0, \dots, K]}$ (Fig.2). Finally, we define a new feature vector $\gamma(r^K)$ associated to each $r^K \in \mathcal{P}^K$ as the concatenation of all the visual features θ assigned to supervoxels of $\mathcal{E}(r^K)$. Concatenating all visual features across \mathcal{M} reaches a powerful multi-scale description of finest scale supervoxels.

A tissue classification based on supervoxels of scale K is carried out via standard RF [1]. RF consists of T decorrelated trees made of internal nodes splitting data according to binary tests h and leaf nodes which reach all together a final data partition. At each internal node, the split sends supervoxels r^K to left and right children nodes. For this task, the associated binary test focuses on a randomly selected subset $\hat{\gamma}(r^K)$ of

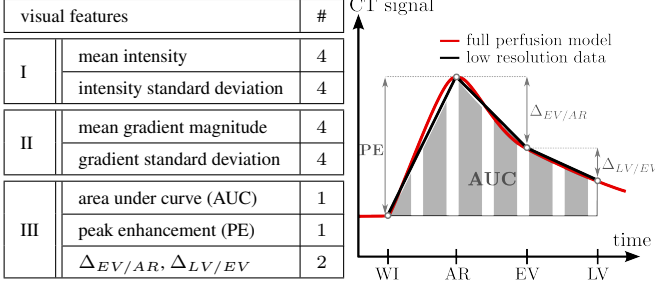


Fig. 3. Multi-phase supervoxel-based visual features involved into random forest with illustration of dynamic (III) ones.

features $\gamma(r^K)$ and halves the input supervoxel set following:

$$h(r^K, \gamma(r^K)) = \begin{cases} \text{true, if } \tau_{low} < \hat{\gamma}(r^K) < \tau_{up} \\ \text{false, otherwise} \end{cases} \quad (1)$$

Internal node parameters ($\{\tau_{low}, \tau_{up}, \hat{\gamma}(r^K)\}$) are optimized via information gain maximization with respect to $\mathcal{S} = \{r_i^K, c(r_i^K)\}_{i \in \{1, \dots, K_u\}}$, a set of K_u manually labeled training supervoxels $\{r_i^K\}$ where each r_i^K receives an associated ground truth label $c(r_i^K) \in \{c_1, \dots, c_N\}$. After optimization, each leaf node l_t of the t^{th} tree receives a partition \mathcal{S}_{l_t} of training data \mathcal{S} and produces an entire class probability distribution: $p_{l_t}(c_l | \mathcal{S}) \forall l \in \{1, \dots, N\}$. To predict label $c(r^K)$ of a given test supervoxel $r^K \in \mathcal{P}^K \setminus \mathcal{S}$ with associated features $\gamma(r^K)$, r^K is injected into each optimized tree which makes it reach a leaf node l_t per tree following successive split rules. For each label c_l , we get:

$$p(c(r^K) = c_l) = \frac{1}{T} \sum_{t=1}^T \frac{|\{r_i^K, c(r_i^K)\} \in \mathcal{S}_{l_t} \mid c(r_i^K) = c_l\}|}{|\mathcal{S}_{l_t}|} \quad (2)$$

The final prediction of $c(r^K)$ corresponds to the label c_l maximizing $p(c(r^K) = c_l)$.

2.4. Multi-phase supervoxel-based visual features

Visual features $\theta(r^k)$ assigned to each 3D supervoxel r^k of \mathcal{M} are divided into three categories (Fig.3): intensity (I), gradient (II) and dynamic (III) information [6]. The two first introduce spatial characteristics in terms of intensity and gradient magnitude at the supervoxel spatial extent. In both cases and for each phase, mean and standard deviation values are computed among all voxels of r^k . The third category exploits DCE data by combining area under curve (AUC), peak enhancement (PE) as well as other inter-phase intensity differences $\Delta_{EV/AR}$ and $\Delta_{LV/EV}$ averaged among voxels of r^k .

When building \mathcal{M} , supervoxels r^k with less than K_v voxels are no longer split. Under this threshold, intensity and gradient standard deviations used as features would become meaningless. These smallest supervoxels are simply duplicated from scale $k+1$ to scale K to obtain concatenated feature vectors $\gamma(r^K)$ with similar length at the finest scale.

methods	optimized sSLIC-RF [6]		hSLIC-RF
resolution	shared	e_i -dependent	multi-scale
DICE _{activ}	76.5 ± 10.1	78.7 ± 9.18	80.4 ± 8.81
DICE _{necro}	85.3 ± 12.5	86.9 ± 9.51	86.9 ± 10.5
DICE _{prom}	94.3 ± 4.12	94.9 ± 3.85	95.5 ± 3.56
DICE _{tumor}	88.9 ± 8.51	89.4 ± 6.12	91.0 ± 6.99

Table 1. Quantitative comparisons of optimized single-scale supervoxel-based random forest (RF) sSLIC-RF [6] and our hierarchical multi-scale supervoxel-based RF hSLIC-RF through DICE coefficients averaged over the whole database.

3. RESULTS

3.1. Evaluation on clinical data

Experiments focus on data collected from 8 examinations $\{e_1, \dots, e_8\}$ stemming from patients with HCC. Each examination results in a set of DCE-CT scans. For each e_i , 6 equally reparted 2D axial slices have been selected in I_{EV} to cover the tumor extent and labeled by 4 experts in hepatodigestive surgery to reach ground truth (GT) masks delimitating parenchyma, active and necrotic tissues ($N=3$). It results in a database of 48 slices with associated fused GT masks obtained by keeping common expert annotations only. This database slightly differs from the one used in [6].

Our goal is to assess the proposed hierarchical multi-scale supervoxel-based RF (hSLIC-RF) and to provide comparisons with the single-scale supervoxel-based RF of [6] (sSLIC-RF). Each e_i is processed independently. Resulting HCC tumor segmentation results are assessed via DICE coefficients between obtained and fused GT masks for parenchyma, active, necrotic and tumoral tissues. Each forest contains $T=100$ trees. To be less sensitive to variability due to both RF random aspects and random training selection, results for each e_i are averaged over 10 realizations.

Since sSLIC-RF results can widely vary from one supervoxel resolution to another, we tested sSLIC-RF on each e_i with various resolutions from 6mm^3 to 20mm^3 with a step of 2. We identified the optimal resolution for each e_i based on the best averaged DICE among the 3 tissues. Although not feasible in clinical practice, such an exhaustive search allows to get the best results reachable by sSLIC-RF. Results from sSLIC-RF are provided with optimal shared (same resolution $\forall e_i$) and e_i -dependent resolutions. hSLIC-RF is performed with $K=9$, $K_s=6$ and $K_v=100$.

Training data \mathcal{S} resp. extracted from \mathcal{P}^K and single-scale supervoxels for h/sSLIC-RF are obtained as follows. We identify supervoxels whose intersection with annotated slices has a predominant GT label (consensus over at least 95% of voxels). This label is assigned to the supervoxel as its own GT label. By this way, we simulate a balanced interaction where the practitioner only labels 15 instances of each class ($K_u=45$). In average, it represents 0.055% of Ω_l for hSLIC-RF.

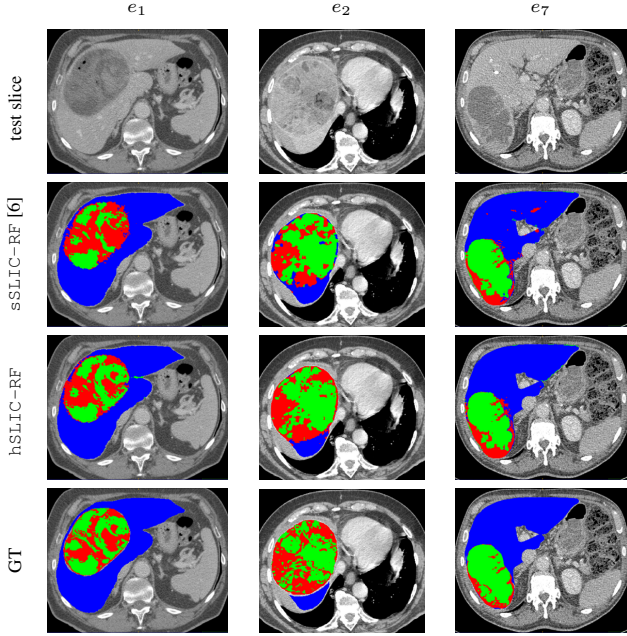


Fig. 4. HCC liver tumor segmentation results via single-scale supervoxel-based RF $sSLIC-RF$ [6] and our hierarchical multi-scale supervoxel-based RF $hSLIC-RF$ with expert GT annotations fused with STAPLE [11] to allow a visual comparison with dense GT masks. Parenchyma, active and necrotic areas are respectively in blue, red and green.

3.2. Discussion

We present in Tab. I a comparative assessment of $hSLIC-RF$ and $sSLIC-RF$. It reveals globally better results using $hSLIC-RF$ which achieves highest DICE for parenchyma (95.5), active (80.4) and tumoral (91.0) areas. Except for $DICE_{necro}$ whose results are similar, $hSLIC-RF$ outperforms the upper bound reachable by $sSLIC-RF$ with gains on $DICE_{\{prcm,activ,tumor\}}$ respectively about 0.6, 1.7 and 1.6. Compared to $sSLIC-RF$ with shared resolution, gains on $DICE_{\{prcm,activ,necro,tumor\}}$ are respectively 1.2, 3.9, 1.6 and 2.1 combined with reduced standard deviations. It confirms the benefits of our adaptive data sampling scheme. Moreover, comparisons between $sSLIC-RF$ with shared and e_i -dependent resolutions show the significant impact of scale selection in the single-scale context. According to visual results and expert GT annotations fused with STAPLE [11] (Fig. 4), $hSLIC-RF$ reaches a better segmentation accuracy than $sSLIC-RF$. We notice the ability of $hSLIC-RF$ to get stronger spatial regularization inherited from the capacity of multi-scale SLIC supervoxels to adhere to image boundaries.

4. CONCLUSION

In this work, we extend supervoxel-based random forest classification from single to multi-scale in the context of HCC

liver tumor segmentation. Since intrinsic tissue properties may emerge at different scales for different tissues, we propose to embed multi-scale visual features in a hierarchical multi-scale tree resulting from recursive supervoxel decomposition. This approach provides a powerful multi-scale description of leaf supervoxels on which a standard random forest is applied to accurately discriminate parenchyma, active and necrotic tissues. We demonstrate the effectiveness of our method on challenging clinical data and will further explore this scheme in the context of multi-examination training.

5. REFERENCES

- [1] L. Breiman, "Random Forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [2] Zikic D. et al, "Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel MR," in *Medical Image Computing and Computer-Assisted Intervention*, pp. 369–376. 2012.
- [3] A. Montillo, J. Shotton, J. Winn, J. E. Iglesias, D. Metaxas, and A. Criminisi, "Entangled decision forests and their application for semantic segmentation of CT images," in *Information Processing in Medical Imaging*, 2011, pp. 184–196.
- [4] S. Bauer, L.-P. Nolte, and M. Reyes, "Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization," in *Medical Image Computing and Computer-Assisted Intervention*, pp. 354–361. 2011.
- [5] W. Wu, A. YC Chen, L. Zhao, and J. J Corso, "Brain tumor detection and segmentation in a CRF framework with pixel-pairwise affinity and superpixel-level features," *International journal of computer assisted radiology and surgery*, vol. 9, no. 2, pp. 241–253, 2014.
- [6] P.-H. Conze, F. Rousseau, V. Noblet, F. Heitz, R. Memeo, and P. Pessaux, "Semi-automatic liver tumor segmentation in dynamic contrast-enhanced CT scans using random forests and supervoxels," in *Machine Learning in Medical Imaging*, vol. 9352, pp. 212–219. 2015.
- [7] L. Peter, O. Pauly, P. Chatelain, D. Mateus, and N. Navab, "Scale-adaptive forest training via an efficient feature sampling scheme," in *Medical Image Computing and Computer-Assisted Intervention*, pp. 637–644. 2015.
- [8] E. Geremia, B. H Menze, and N. Ayache, "Spatially adaptive random forests," in *IEEE International Symposium on Biomedical Imaging*, 2013, pp. 1344–1347.
- [9] M. Ronot and V. Vilgrain, "Hepatocellular carcinoma: Diagnostic criteria by imaging techniques," *Best Practice & Research Clinical Gastro-enterology*, vol. 28, no. 5, 2014.
- [10] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, pp. 2274–2282, 2012.
- [11] S. K Warfield, K. H Zou, and W. M Wells, "Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation," *IEEE Transactions on Medical Imaging*, vol. 23, no. 7, pp. 903–921, 2004.